心理表征的可视化途径：基于噪音的反向相关图像分类技术*

侯春娜[1]**　　刘志军[2]
[1](东北师范大学心理学院，长春 130024)
[2](长春理工大学社会学系，长春 130022)

摘要： 社会心理学对图像的心理表征研究一直难以将心理活动的内容准确刻画出来。近十年来出现了一种新心理物理学方法——“反向相关图像分类技术”，该技术假定观察者的反应与视觉噪音存在相关关系，且反应是依照观察者的社会判断标准进行而非随机做出；通过对其做出反应的相应噪音模式的足够次数的权重计算与视觉代码显现，从而将观察者内在的评估特点可视化。该技术已在特质研究、种族和群际偏见等领域取得了一些成果，但是未来仍需解决实验次数过多，分离混杂的噪音以及被试的表现等问题，才能获得更为真实的心理表征。

关键词：面孔 心理表征 反向相关图像分类技术
分类号：B841

Visualization of Mental Representation: Noise-based Reverse Correlation Image Classification Technology

HOU Chun-Na[1]　　LIU Zhi-Jun[2]
[1](1 School of psychology, Northeast Normal University, Changchun 130024, China)
[2](Department of Sociology, Changchun University of Science and Technology, Changchun 130022, China)

Abstract:
Studies of the mental representation of images in social psychology have encountered difficulty in accurately portraying psychological activity. Over the past decade, reverse correlation image classification has emerged as a new psychophysical method that assumes there is a relationship between an observer’s response and visual noise, and that the response is based on the observer’s social judgment criteria, and are not random. Performing a sufficient number of weight calculations on the corresponding noise patterns of the observer’s reaction allows us to visualize the intrinsic evaluation characteristics of the observer. The use of reverse correlation image classification technology has achieved some results in the areas of trait research, ethnicity, and intergroup bias. In the future, however, it is necessary to solve the

* 本文系教育部人文社会科学研究青年基金(16YJC90009)与中央高校基本科研业务费专项资金资助(2412017QD030)的研究成果之一。
** 通信作者：侯春娜， E-mail: houcn359@nenu.edu.cn

problems of excessive experimental trials, separation of mixed noise, and subjects' performance, in order to achieve more realistic mental representations.
Keywords: face, mental representation, reverse correlation image classification technology

引言

表征(representation) 也称心理表征和知识表征，其含义是外部事物在心理活动中的内部再现。一直以来，心理表征的研究多集中于认知心理学的研究，多关注知识概念层面的研究。在社会知觉过程中，人们需要快速的对面孔特征，如种族、性别、年龄、可信性等社会属性做出推断(Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015)，由于推断的属性不能直接被观察到，所以认知系统必须将感知输入的视觉信息与心理表征匹配来实现这一点。然而受传统心理物理法的局限，人们难以将面孔内容信息在心理活动中的表现和记载的方式直观化呈现，因而对心理表征的观测一直存在困难。

反向相关图像分类技术(Reverse Correlation Image Classification，RCIC)是一种纯粹基于数据驱动的心理物理学方法。此种技术被证实能够将心理表征的内容直观可视化，揭示人们的内在表征和决策策略(Dotsch, Wigboldus, Langner, & van Knippenberg, 2008)。经过十余年的发展，RCIC 技术已成为揭示人们心理表征，进行高级知觉或认知过程研究的一种“有力工具”(Ponsot, Arias, & Aucouturier, 2018)。基于此，本研究对 RCIC 的的研究进展进行了剖析，以期为本土社会知觉心理表征的研究提供一种可视化的心理物理学方法指导，为拓展未来社会心理学的研究提供一种从心理表征内容出发新的研究途径。

1. RCIC 的起源与数理基础

RCIC 发展自反向相关技术，后者在过去四十年中被广泛应用于视觉和听觉的知觉过程研究。反向相关技术本质上遵从信号检测理论，但是与通过被试甄别刺激信号的标准实验程序相反。传统的信号检测范式中，反应依赖于对刺激属性的有意义操作，两者间的关系是由固定刺激与反应的属性所决定的，即同样刺激可诱发出不同反应。但在反向相关技术中，其假定由于不同刺激的推断模式导致了不同的反应，因而刺激不再固定不变而是随机变化的，由被试自行决定信号是否存在刺激中。这种技术通过对产生虚惊的噪音场的收集和组合，以推断该过程中用于执行特定任务的策略。其中，术语“反向(reverse)”即是指刺激和反应之间统计学关系的反转之意。RCIC 能够通过刺激噪音参数的随机变化及对刺激判断中每个刺激噪音模式相对权重的估计而生成直观的表征内容视觉图像代码。

RCIC 的数理基础可追溯至数学中的相关。在视觉研究中，依照视觉刺激信号从视网膜的输出神经元经由外侧膝状核(lateral geniculate nucleus)，传递至初级视觉皮层(primary visual cortex)的过程，刺激与神经反应两种信号之间随时间滞后的相关关系，可用积分公式表现为：

$$Corr[g,h](t)=\int_{-\infty}^{\infty}dt'g(t')h(t+t')$$

其中 g 和 h 分别表示刺激与神经反应两种信号，t 表示时间。对于 t 条件而言，这两个无限维度或有限维向量的相关可进一步演化为：

$$Corr(g,h)_j=\sum_{i=-\infty}^{\infty}g_i h_{i+j}$$

通过维纳-霍普夫(Wiener-Hopf)积分方程，可以实现连续时间平稳过程的方差估计误差达到极小，以此满足最优滤波器脉冲响应函数的必要条件。基于此，方程假定存在两个时间序列刺激 x_i 和神经反应 y_i，将 y 建模为 x 的滤波版本，则上述相关公式就近似表达为：

$$y_i \approx \sum_j h_j x_{i-j}$$

通过寻找 *h* 的最优值，也即内核(Kernel)，可以求解这个方程(Ahumada & Beard, 1998)。RCIC 就是用于发现线性系统的内核，并将视觉输入和输出进行关联的心理物理学方法。通过将高斯白噪音应用于刺激之上，试图发现视觉系统的线性系统内核。即将高斯白噪音 x 带入方程，则 Wiener-Hopf 方程的内核的解便演化为：

$$h_k = \frac{C_k^{xy}}{C_0^{xx}} \text{(注：这里的} C_{k-j}^{xx} = C_0^{xx}\delta_{kj}\text{)}$$

为了实现这一点，其具体做法就是在传统的信号检测实验中将外部滤波信号(视觉噪声)添加至每次试验中的原始刺激上。通过大量数目的观测试验并计算出人们迫选出的噪音像素（并非原始刺激）集合，将之进行权重后显示为图像便实现了内核的可视化，这个直观的内核有时也被称为分类图像(Classification Image, CI; Saegusa, Yamaoka, & Watanabe, 2015)。由于这个内核图像是人们对外部滤波信号的反应，蕴含着被试的社会判断标准（特质评估），将其叠加在原始实验刺激上即能够生成反映观察者期望效应的表征图像(Dotsch & Todorov, 2012)。

不仅如此，通过分析 CI 图像像素间的内在关系，RCIC 还可以推测出观察者如何做出决策。它的数学原理是基于随机场理论(Random Field Theory, RFT)实现的。随机场理论具有两个核心要素：位置(site)与相空间(phase space)。依照 RFT 的观点，当给每一个位置中按照某种分布随机赋予相空间的一个值之后，此全体就叫做随机场。Adler 和 Hasofer (1976)推导出固定高斯随机场的预期欧拉特征(Euler Characteristic, EC) 的偏差度，可以作为激活区域数量的估计量。EC 本质上计量了平滑高斯随机场中高于足够高阈值的簇数量，通过将超过固定阈值点的像素进行簇聚（聚类分析），在空间上构成的视觉代码即反映了人们的推断依据。这个视觉代码叠加至基本刺激（如面孔）上，就得到了人们进行特质评估的诊断区域(Dotsch & Todorov, 2012)。

综上，RCIC 的技术原理可以概括为：RCIC 假定观察者的反应与视觉噪音存在相关关系，且反应是依照观察者的社会判断标准进行而非随机做出；通过对其做出反应的相应噪音模式的足够次数的权重计算与视觉代码显现，从而将观察者内在的评估特点可视化；进而通过数学逻辑运算推测出观察者的决策规律。学者认为 RCIC 可以作为研究心理表征的切入点(Brinkman, Todorov, & Dotsch, 2017)。RCIC 生成表征图像与聚类分析的具体流程见图 1 所示。

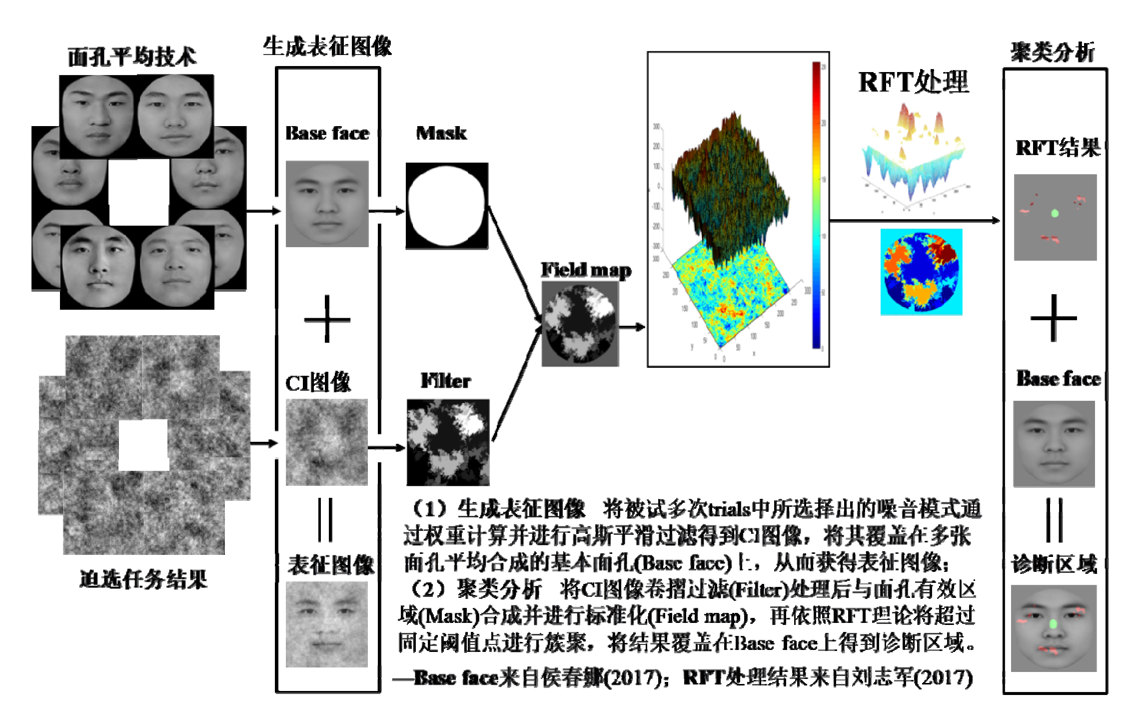


图 1　RCIC 生成表征图像与聚类分析示意图

2．研究范式的变革

反向相关的研究范式主要包含两种过程：随机噪音刺激的生成过程与心理表征图像的生成过程。伴随着技术原理的发展及对心理表征效果的考察，该范式在使用噪音的类型及任务模式上的变化大致经历了三个阶段。

2.1　纯粹随机噪音研究范式

视觉系统研究认为在早期视觉信息采集与高级的认知加工之间，必定存在着一个中间过程，这一过程涉及对视觉场景全貌的表征构建。受到采用反向相关技术模拟被试在解决空间立体任务时形成表面模板(surface template)的启示，学者发现将被试在实验中所选择的全部噪音模式进行平均可以得到一个 CI 图像，这个分类图像映射着被试对随机噪音刺激的反应变化，可被视为内在“模版”(Saegusa et al., 2015)。为了避免心理表征图像是刺激中残存信号所得的伪像，Gosselin，Bacon 和 Mamassian(2004)借助反向相关技术设计了一种纯粹由任务决定的图像表征实验范式，他们首先在白色背景上随机生成了 700 个黑点的纹理图像，然后通过立体镜成像为立体图。各个纹理图像的位置和深度完全以随机的方式确定，因而立体图中并不包含任何信号成分。实验任务是要求被试报告是否在整个屏幕区域侦测到一个立体的“+”或“–”，尽管事实上并不真实存在这样的图像。通过整合所有反应，研究者证实被试能够在随机噪音中 “拉出”一个清晰的 “+” 或“–”图像，从而呈现出被试对符号的内在表征估计。

由于整个实验过程并不包含结构性符号元素，因此这种范式能够跨越较为早期的低水平的视觉加工阶段而又不影响视觉识别等其他的高级加工过程。借助反向相关技术，学者首次实现了对心理表征的模拟与描绘，证实了视觉信号加工与高层次认知加工间存在着心理表征这样的中间过程(Gosselin et al., 2004)。Gosselin 等人的研究使探索心理表征内容成为了可能。

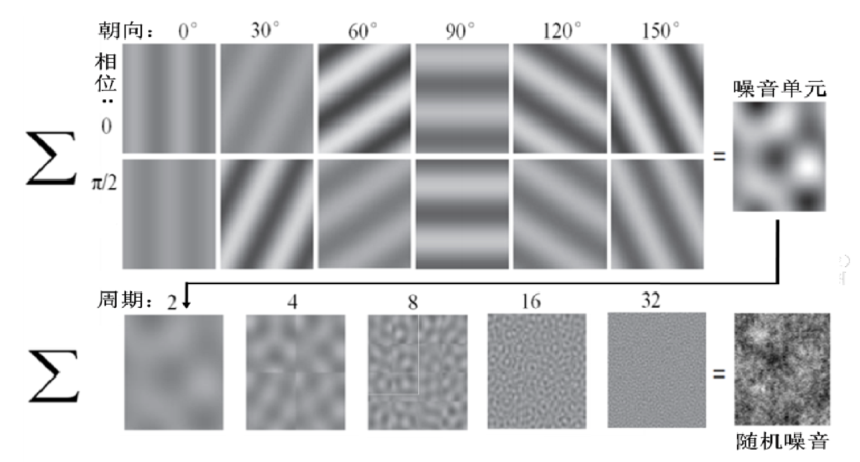
值得注意的是，上述任务中的实验刺激所叠加的噪音采用了高斯白噪音，通过对所有高斯白噪音像素强度加权求和从而将信息投射到线性空间上。有研究者对此提出了质疑，认为

这种线性评估仅被证实在低层次的视觉神经生理学和心理物理学研究中富有成效，但对高级社会认知活动的作用仍不得而知。而且，这种范式中并没有对基本图像（空白背景）进行约束，造成刺激空间过大，研究中需要巨大的试验次数。以 Gosselin 等人(2004)的 20000 次实验为例，这极大增加了被试的负担。

2.2　正弦随机噪音研究范式

Mangini 和 Biederman (2004)采用随机变化的正弦噪音替代高斯白噪音，相比之下正弦曲线更具优势。这是因为：首先，正弦曲线更接近早期皮层视觉区域的优选刺激。其次，由于噪声谱(noise spectrum)类似于信号分布，观察者更可能将噪声解释为刺激变化。虽然这样的功率谱(power spectrum)可以通过对白噪声进行滤波来实现，但是正弦噪声的功率谱比白噪声要少四倍的参数。更为重要的是，在少量的实验次数中（例如少于 1000 次），正弦噪声收敛要较比白噪声估计的更为准确，这无疑有益于降低实验次数，减轻被试的负担。

基于此，Mangini 等人(2004)将反向相关研究的白噪音改进为正弦噪音，他们以 2 个周期截断的正弦曲线图像片段在 6 个不同朝向（0°，30°，60°，90°，120°和 150°）×2 个相位（0，π/2）的变化为基础。将这 12 种图像进行融合得到随机噪音单位，之后将其依照 5 种空间尺度进行排列（每个图形 2，4，8，16 和 32 个周期），最后将一系列对比度随机的图像相叠加且对比度随机化，形成了包含 4092 个正弦函数参数的随机噪音图像（见图 2），以此作为原始的随机场素材。



生成随机噪音：（1）以2个周期截断的正弦曲线图像片段在6个不同朝向（0°，30°，60°，90°，120°和150°）×2个相位（0，π/2）的12种图像变化融合成噪音单元；（2）将之依照5种空间尺度（2，4，8，16和32个周期）进行排列，形成包含4092个正弦函数参数的随机噪音图像。　—来自Mangini, &Biederman (2004)

图 2 随机噪音生成示意图

正弦噪音的周期参数设置很好地兼顾了不同层次的社会认知需要，较低频率能够传递性别和表情信息，而复杂的社会信息（如个性化）需要较高的频率传递。这表明正弦曲线（或 Gabor）不仅与早期皮质阶段的调谐很好地匹配，而且还能灵活有效地反映出人们依据不同任务对不同面孔位置和空间尺度信息的加工结果。

2.3　双图片迫选任务研究范式

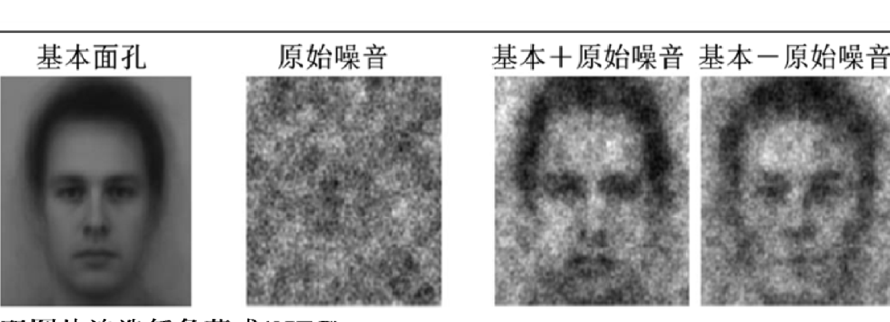
虽然 Mangini 等人(2004)的研究范式能将被试不同决策的随机像素矩阵转化为直观的且具有显著差异的分类图像，但此研究范式应用到建构社会认知模型的程度似乎有限。

首先，这种范式中基本面孔是由两种不同属性的图像合成的，这两种图像代表着两种不

同的目标归类，即在呈现刺激之前便已预设了两种不同的归类。但是在构建社会维度模型时，研究者却假定这些图片并不实际表征两种归类；相反，这些图片的归类恰恰是研究者们力图着手发现的，因而这是一种潜在的矛盾。其次，这个范式最初是以 1-4 级选项迫选任务的形式进行的。在生成心理表征图像过程中，却仅筛选“极有可能”这种较为确信的反应进行加权计算，而不考虑可能性较低的判断，实验效率相对较低。

基于此，Dotsch 等人(2008)采用双图片迫选任务范式 (Two-Image Forced Choice Task，2IFC)对前人的任务模式进行了改进。他们依然选用正弦噪音作为随机噪音，但基本面孔则选择面孔数据库中的平均面孔作为基本图像。这有效消除了实验者预设归类的可能，从而确保分类图像是被试自身心理表征的结果。此外，这种平均面孔由于其轮廓较为模糊，会使面部的其他特征（如嘴巴、鼻子、眼睛等）成为焦点，非常适合叠加随机噪音。

更为重要的是，他们将等级评定的任务模式调整为双图片迫选任务形式。这种改进提升了对实验数据的利用情况：一方面，越多的实验数据则愈加有益于成像效果的清晰度；另一方面，通过平均所有被选择的噪音模式及未被选择的噪音模式，仅一次实验研究就可以同时获得被试（极）期望的分类图像(CI)与（极）不期望的分类图像(antiCI)两种相反的分类图像，从而有效提高了实验的效率。再者，双图片是由随机噪音模式及相反的负性噪音模式分别覆盖在基本面孔上生成，在呈现效果上，这与照片的底片或医学中的正片与负片相似。不仅具有较好的视觉对比效果，而且相反的随机噪音模式还能将所呈现的两个图片间的差异最大化。从而可以将刺激对的数量最小化，缩减了实验次数。2IFC 的双图片生成过程及实验任务呈现形式见图 3。



双图片迫选任务范式(2IFC)
（1）双图片刺激生成 将RCIC生成的原始噪音以及与之相应负性噪音，分别叠加至基本面孔(base face)生成两种噪音模式相反的刺激图像；
（2）双图片迫选任务 每次试验中将两种相反噪音模式的刺激图像并排呈现，要求被试选择出最接近目标类别的面孔。
——来自Brinkman, Todorov, &Dotsch(2017)

图 3 双图片迫选任务研究范式示意图

Dotsch 等人(2008)将实验范式区分为生成随机噪音刺激、执行双图片迫选任务、产生分类图像及执行像素聚类分析四个组成部分。学者认为此种范式将面孔刺激生成过程与特质形成过程相分离，能够从机制上客观揭示出面孔表征直接导致的特质推断过程(Ratner, Dotsch, Wigboldus, van Knippenberg, & Amodio, 2014)。目前，已经广泛应用于社会知觉的相关研究中(Karremans, Dotsch, & Corneille, 2011; Paulus, Rohr, Dotsch, & Wentura, 2016)。

3．RCIC 的技术优势

3.1 方法原理优势

由于表征是内在的心理建构，传统上人们只能依赖于间接方法来推测它们的内容，为了探究复杂的社会刺激如何被识别，人们首先要确定知觉者在判断中会使用刺激的哪些特征以

及如何对这些变化线索进行组合。更具体地说，在被试进行刺激判断之前，研究者需要掌握复杂刺激的不同成份，特征或区域的权重，通过测量和检验刺激的加权模式，才能推测影响个体社会知觉的心理表征内容及决策策略。

传统数据驱动技术的不足在于，其依赖于构建模型的刺激的人为操纵。通过系统操纵刺激的特征来考察人们的社会判断依据。然而一旦对社会属性非常重要的结构特征在一系列判断中不发生变化，则这种方法就无法识别出这些重要的特征(Todorov et al., 2015)。此外，人们在定义刺激特征上还存在不足，一些特征由于无法标识，研究者和知觉者可能都无法意识到这些特征对社会知觉的影响。

相比之下，RCIC 作为一种新的纯粹数据驱动技术则在方法原理上具有更强的优势。受益于工程和神经生理学系统识别技术（如 fMRI），RCIC 同时包含着心理物理学，实验心理学和计算机科学的原理。这种方法对刺激与反应之间关系的探究依赖于实验中每次具体的实验试次分析，而非凭借对刺激特征的操纵。RCIC 通过大量次数的实验，对被试反应的噪音模型进行归类，将选择相同的噪音样本进行平均便可得到分类图像。将经验的分类图像与不同模型假设下推导出的心理表征进行比较，从而辨别何种理论模型更适合于解释被试在任务中的决策行为。这种技术背后的一般原则是：在多次实验中，通过对一个刺激不同部分的随机干扰（“噪音”）及被试相应反应之间的统计关系进行分析直接洞察被试在知觉任务中所使用的心理表征，从而确定哪些知觉决策策略调节任务绩效。其目的在于以尽可能小的偏差来识别高维度变量（如面孔图像）与行为（如感知决定）之间的定量关系。

RCIC 虽然受到特定实验范式中所使用判断和刺激类型的限制，但它对反映社会认知的刺激信息所实行的可视化手段却是无约束的。此种方法通过要求被试对叠加随机噪音的基本图像进行判断，由于随机噪音在像素水平上扭曲了基本图像（如面孔），因此形成的刺激变化要较传统心理物理法受到的约束更小。这种方法的最终目的是为了计算出分类图像，从而将驱动社会判断的心理特征可视化。需要注意的是 RCIC 的可视化效果与眼动追踪研究有着根本的不同。眼动追踪技术识别的只是呈现完整图像时，被试视网膜中央凹所形成的部分图像；而 RCIC 识别则为分类提供潜在信息的整个刺激可视区域。

3.2 生态效度优势

RCIC 另一个优势是生态效度较高。当信号特征未知或想要考察分类表征时，由于没有任何先验假定图像的偏差，使得这种范式不易受社会期望或需求特征的影响。RCIC 也继而成为挖掘高级心理过程中自发使用信息的一种理想方法。

虽然 RCIC 在生成心理表征图像中需要选择一个恒定的原始刺激作为基本图像，但由于这种技术的原理是藉由反应与视觉噪音之间相关关系来执行，原始刺激并不会影响 RCIC 的生态效度。因而，无论选择人类的面孔或躯体作为基本图像(Johnson, Iida, & Tassinary, 2012; Lick, Carpinella, Preciado, Spunt, & Johnson, 2013; Nunnari & Heloir, 2017)，还是选择以非人类图像如狒狒头像(Martin-Malivel, Mangini, Fagot, & Biederman, 2010)及蜘蛛图像为基本图像(Young, 2014)都不会对最终的分类表征图像产生影响。这是因为叠加的随机噪音在视觉效果上完全改变了基本图像的外观。这样便能对基本图像进行多种解释，这也由此创造了依照心理表征中的特征而非基本图像来“塑造”结果图像的机会。学者认为噪音的随机化为 RCIC 的生态效度提供了保障(Paulus et al., 2016)。

以人类面孔为基本图像为例。从种族角度看，目前基本面孔的选择既有高加索白人男性面孔(Dotsch & Todorov, 2012)，黄色人种的印度男性面孔(Dunham, Srinivasan, Dotsch, &

Barner, 2014)也有白人与黑人的混合男性面孔(Brown-Iannuzzi, Dotsch, Cooley, & Payne, 2017)。Dotsch 等人(2008)的研究显示，RCIC 能够将不同情境中被试的想法视觉化，即使以白人面孔为基础面孔，也可以生成摩洛哥人与中国人两种非白人面孔的心理表征(见图 4 左)。而 Martin-Malivel 与同事(2010)年以狒狒头像为基本图像，对狒狒的信息分类问题所进行的比较研究，进一步将 RCIC 的研究对象的范围从人类拓展至动物群体。这些研究均表明，RCIC 方法本身是可靠的，其能够为不同对象的心理表征提供证据。即使在更加自然的情境中，RCIC 也可以同样适用 (Paulus et al., 2016)。

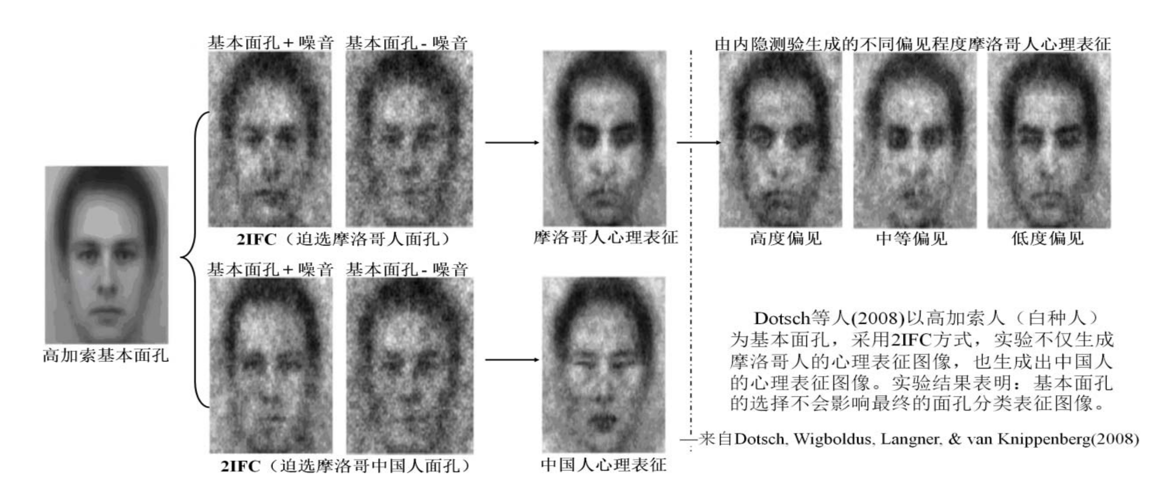


图 4 基于相同基本面孔获得的摩洛哥人与中国人的面孔表征

4. RCIC 的研究应用

特质评估 Oosterhof 和 Todorov(2008)采用数据驱动技术构建面孔模型，发现人们在形成第一印象中，主要以可信性和支配性两种维度对他人进行特质评估。但此特质归因模型仍需证实。得益于 RCIC 无需任何先验假定的技术优势，Dotsch 和 Todorov (2012)不仅再次证实了特质归因模型的有效性，还发现可信性高的面孔表征形象显得更加女性化且伴有微笑表情；而可信性低的面孔表征形象更具男性特征且伴有愤怒表情，从而为情绪泛化解释观点及可信性的性别差异现象均提供了证据支持。后续，Imhoff 等人(2013)使用 RCIC 对特质归因模型中两维度的关系展开了进一步探索，两种特质会表现出此高彼低的相互补偿关系(Imhoff, Woelki, Hanke, & Dotsch, 2013)。不过近来 RCIC 的相关研究表明，人们无法使用这两个维度形成自我形象(van Driel, 2017)，这为特质研究的发展提供了新的思考点。

此外，Nunnari 和 Heloir (2017)基于躯体基本图像的 RCIC 范式，以蕴含着 14 种躯体属性的虚拟人物角色（包含面孔与躯体的人物形象）为基本图像，考察了 3 种人格特质(可信性、支配性、宜人性）与躯体属性之间的关系，并尝试构建一种特质与躯体属性系统(traits-to-attributes system)的线性关系模型，这为特质研究做出了新的尝试。

群际关系 Ratner 等人(2014)对群际关系中的视觉表征作用进行了考察。借助 RCIC，研究者获得了内群成员与外群成员的典型视觉表征形象，并发现人们对内群成员分类图像的特质评估要比外群分类图像更加积极，证实视觉表征是传递群体偏见的一种强大机制。这一发现也为基于社会认同理论框架来解释群际关系提供了重要的视觉表征证据。不仅如此，即使普通的面孔表情，也会受到群体成员身份的影响。Paulus 等人(2016)通过 RCIC，发现内群微笑面孔的视觉表征形象传递出更多的社会意义（仁慈），从而表现出内群偏私现象。

但近来的 RCIC 研究也证实人们可以既存在内群偏私现象也存在外群偏私，这可能对社

会认同理论的解释力提出了挑战。一项针对印度儿童宗教表征与种姓表征评估的结果表明，在种姓上比达利特儿童对高社会地位婆罗门分类图像具有更积极的评价，但同时在宗教上又强烈表现出对下层阶层穆斯林教而非上层阶层印度教的评价偏好 (Dunham et al., 2014)。

偏见 Dotsch 等人 (2008) 利用 RCIC 对高级社会认知中的偏见进行了探索。已有研究显示，偏见会导致对种族外群的认知、情感和行为产生偏差(Fiske, 2015)。这意味着由于某种预设动机或特定知识的影响，人们在观察外群面孔刺激时可能将造成外群的面孔表征形象出现扭曲。基于这种推测，Dotsch 等人 (2008)以一个在荷兰被高度污名化的移民群体（摩洛哥人）为目标群体，考察了人们对外群成员的心理表征是否受到潜在偏见的影响。分类图像结果显示，被试对摩洛哥人内隐偏见越高，越倾向将其原型表征为一种在客观判断上看起来更不可信且更像罪犯的形象，即越具偏见的个体对外群心理表征的扭曲程度越大，其产生的外群分类形象会越加消极。这一结果揭示外群体的心理表征形象可能与偏见水平有着密切的联系（见图 4 右）。

此外，在性别研究中，学者通过基于躯体为基本图像的 RCIC 技术，将人们内在的两性躯体表征形象进行了直观的刻画。图像处理结果发现，相比于男性，人们对女性存在着不同程度的性别偏见(Johnson et al., 2012; Lick et al., 2013)。

心理治疗 研究表明蜘蛛恐怖症患者对蜘蛛存在异常认知，他们会更迅速地注意到蜘蛛图像，具有更强的愤怒情绪体验及更高的焦虑状态。Young(2014)以基于蜘蛛为基本面孔的 RCIC 技术对蜘蛛恐怖患者的蜘蛛心理表征形象进行了考察评估，发现他们形成的典型蜘蛛心理表征更具威胁性，更让人觉得恐惧。从而证实蜘蛛恐惧症患者对蜘蛛的恐惧可能与感知恐惧相关刺激的异常表征有关，这为恐惧症未来的介入与治疗指出了新的方向。

此外，有学者使用 RCIC 对精神分裂症的面孔情绪认知缺陷问题进行了探索，发现相比于正常人，精神分裂症患者过度利用鼻子和嘴巴区域却不能充分利用眼睛区域信息，并且忽略最低空间频率的信息。这些发现为精神分裂症在区分面部情绪过程中使用视觉信息的异常模式提供了直接的证据(Clark, Gosselin, & Goghari, 2013)。

除上述应用研究外，RCIC 还被应用到情绪表达(Jack, Caldara, & Schyns, 2012)，亲密关系的研究中(Gunaydin, & Delong, 2015)。这种技术对种族、年龄、性别、等特征的判断也具有预测性。譬如探究哪种面部特征使人会看起来更像白种人或黑人(Krosch & Amodio, 2014)，或老年人(van Rijsbergen, Jaworska, Rousselet, & Schyns, 2014)，男人或女人(Dotsch,Wigboldus, & van Knippenberg, 2011)。

5．技术局限与前景展望

RCIC 的目的是建立可视化的心理表征内容，虽然这种技术具有较高的生态效度，但其只能够提供真实心理表征的近似值。这是因为由该技术生成的分类图像除了包含真实心理意象外，还有叠加的噪音混杂其中，同时被试的表现也会影响分类图像的生成，如何将这些混杂的因素分离，从而得到真实的心理表征仍然是值得继续深入探究的问题之一。此外对于基本图像的设置也存在不同的声音，有学者通过面孔平均技术所得到的合成面孔作为基础面孔能够降低单独面孔间的差异细节，且具有与单独面孔更多的相似性而被认为更具代表性(Bijvank, 2014)，其他学者则认为 RCIC 技术基于计算机做出的三维面孔作为刺激材料可能更好(Chen, Garrod, Schyns, & Jack, 2017)。

此外，作为一种“分子”心理物理学方法，RCIC 对刺激与反应之间关系的探究依赖于每次具体实验的统计分析，须要大量的实验次数（大多超过 1000 次）才能完成对随机噪音模

式规律的运算(Lick et al., 2013)。这使得此种研究范式在实际操作中面临着一定的困难，如何确定出合理的实验次数是该技术尚需解决的方法学问题。

在过去十年的研究中，RCIC 技术已经被社会心理学家所采用，并已由初期的应用于简单心理过程的研究深入到高级心理过程的诸如偏见、刻板印象和文化层面的研究。它已经被证明是探究心理表征的重要工具，未来将具有更大的应用前景。

首先，研究手段上可与高级认知神经科学相结合。RCIC 可提供直观可视化的心理表征形象，而眼动技术优势则能清晰捕捉到人们的眼动轨迹，未来两种技术的结合可实现优势互补，更有益于高级视觉认知机制的揭示。此外，高级认知活动如何通过心理表征的诊断区域执行社会判断，人们对相应的心理加工机制乃至神经加工机制仍不清晰，而鉴于 RCIC 与 fMRI 相同的数理基础及分别对心理内容与心理过程的研究优势，未来亦可尝试与神经影像的技术结合，更为系统地揭示心理表征加工机制。

其次，应用领域可以继续扩展。如临床心理治疗中的应用：（1）RCIC 亦可应用于特殊儿童（如述情障碍或自闭症儿童）的情绪训练，由于其能够对生成的表征形象进行有效的评估，因而在实际介入中将具有较强的操作性。（2）可应用于恐怖症的系统脱敏治疗。已有研究显示恐怖会导致恐惧相关刺激（如蜘蛛）的心理表征形象出现扭曲，通过引导患者对心理表征形象进行操作，逐渐恢复至正常表征形象，实现对恐怖症的系统脱敏治疗。（3）也可应用于身体形象塑造，通过想象理想的身体形象实现对患有神经性厌食症患者的辅助治疗。此外，在司法领域，由于 RCIC 可以忽略基础图像便可形成被试心理所想的面孔图像，此技术也可应用于目击者对罪犯的肖像临摹。

值得注意的是，作为一种新技术而言，RCIC 仍需要完善。对噪音参数的改进，基本图像从面孔拓展至躯体，甚至非人类图像都是未来值得继续探索的问题，这将有助于 RCIC 范式应用于社会认知更多领域的研究。总之，RCIC 技术能够帮助人们提取出有意义的心理形象，并将心理表征的内容直观可视化，它为高级社会认知加工领域的研究提供了一种新途径。

参考文献：

侯春娜著. (2017). *面孔：群际信任的进化密码*. 北京：科学出版社

刘志军（2017）. *群际认知的面孔补偿效应——基于反向相关图像分类任务的研究*. 吉林大学博士论文.

Adler, R. J., & Hasofer, A. M. (1976). Level crossings for random fields. *The Annals of Probability, 4(1)*,1–12.

Ahumada, A.J., & Beard B.L., (1998). Response classification images in vernier acuity. *Investigative Ophthalmology and Visual Science, 39(4)*,S1109.

Bijvank, M. (2014). Periodic review inventory systems with a service level criterion. *Journal of the Operational Research Society, 65(12)*, 1853–1863.

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: a primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology, 28(1)*, 333–361.

Brown-Iannuzzi, J. L., Dotsch, R., Cooley, E., & Payne, B. K. (2017). The Relationship Between Mental Representations of Welfare Recipients and Attitudes Toward Welfare. *Psychological Science, 28(1)*, 92–103.

Chen, C., Garrod, O., Schyns, P., & Jack, R. (2017). Mapping dynamic conversational facial expressions across cultures. *Journal of Vision, 17(10)*, 834–834.

Clark, C. M., Gosselin, F., & Goghari, V. M. (2013). Aberrant patterns of visual facial information usage in schizophrenia. *Journal of Abnormal Psychology, 122(2)*, 513–519.

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and*

Personality Science, 3(5), 562–571.

Dotsch, R., Wigboldus, D. H., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science, 19(10)*, 978–980.

Dunham, Y., Srinivasan, M., Dotsch, R., & Barner, D. (2014). Religion insulates ingroup evaluations: the development of intergroup attitudes in India. *Developmental Science, 17(2)*, 311–319.

Fiske, S. T. (2015). Intergroup biases: a focus on stereotype content. *Current Opinion in Behavioral Sciences, 3*, 45–50.

Gosselin, F., Bacon, B. A., & Mamassian, P. (2004). Internal surface representations approximated by reverse correlation. *Vision research, 44(21)*, 2515–2520.

Gunaydin, G., & Delong, J. E. (2015). Reverse correlating love: highly passionate women idealize their partner's facial appearance. *Plos One, 10(3)*, e0121094.

Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! visual encoding of stereotype content. *Frontiers in Psychology, 4(386)*,1–8.

Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General, 141*, 19–25.

Johnson, K. L., Iida, M., & Tassinary, L. G. (2012). Person (mis)perception: functionally biased sex categorization of bodies. *Proceedings of the Royal Society B Biological Sciences*, *279*, 4982–4989.

Karremans, J.C.,Dotsch, R., & Corneille,O. (2011). Romantic relationship status biases memory of faces of attractive opposite-sex others: evidence from a reverse-correlation paradigm. *Cognition,121*:422–26

Krosch, A. R., & Amodio, D. M. (2014). Economic scarcity alters the perception of race. *Proceedings of the National Academy of Sciences of the United States of America, 111(25)*, 1–6.

Lick, D. J., Carpinella, C. M., Preciado, M. A., Spunt, R. P., & Johnson, K. L. (2013). Reverse-correlating mental representations of sex-typed bodies: the effect of number of trials on image quality. *Frontiers in Psychology, 4(2)*, 476–484.

Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science, 28(2)*, 209–226.

Martin-Malivel, J., Mangini, M. C., Fagot, J., & Biederman, I. (2010). Do humans and baboons use the same information when categorizing human and baboon faces?. *Psychological Science*, *17(7)*, 599–607.

Nunnari, F., & Heloir, A. (2017). *Generating Virtual Characters from Personality Traits via Reverse Correlation and Linear Programming*. Conference on Autonomous Agents and Multiagent Systems, 1661–1663.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105(32)*, 11087–11092.

Paulus, A., Rohr, M., Dotsch, R., & Wentura, D. (2016). Positive feeling, negative meaning: Visualizing the mental representations of in-group and out-group smiles. *PloS one, 11(3)*:e0151230.

Ponsot, E., Arias, P., & Aucouturier, J. J. (2018). Uncovering mental representations of smiled speech using reverse correlation. *Journal of the Acoustical Society of America*, *143(1)*, 19–24.

Ratner, K. G., Dotsch, R., Wigboldus, D. H., van Knippenberg, A., & Amodio, D. M. (2014). Visualizing minimal ingroup and outgroup faces: implications for impressions, attitudes, and behavior. *Journal of Personality and*

Social Psychology, 106(6), 897–911.
Saegusa, C., Yamaoka, M., & Watanabe, K. (2015). *Seeing faces in noise: Exploring machine and human face detection processes by the reverse correlation method*. Paper presented at the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Siem Reap, Cambodia.
Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (2011). Data-driven methods for modeling social perception. *Social andPersonality Psychology Compass, 5(10)*, 775–791.
Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology, 66(1)*, 519–545
van Driel,S.D.(2017).*Prediction of Self Perception based on Dominance and Trustworthiness by using Reverse Correlation*. (Unpublished master’s thesis). Utrecht University, Netherlands.
Van Rijsbergen, N., Jaworska, K., Rousselet, G. A., & Schyns, P. G. (2014). With Age Comes Representational Wisdom in Social Signals. *Current Biology, 24*, 2792–2796.
Young, A. I. (2014). *Seeing Scary: Predicting Variation in the Scariness of the Mental Representations of Spiders*. (Unpublished doctoral dissertation). The Ohio State University, Ohio State.

作者贡献声明：

侯春娜：提出研究思路，架构研究逻辑，论文写作，修改和审定论文
刘志军，搜集整理文献，修改和审定论文